# DUAL-ROUTE PHONETIC ENCODING: SOME ACOUSTIC EVIDENCE

*S. P. Whiteside and R. A. Varley*

Department of Human Communication Sciences, University of Sheffield, UK.

## ABSTRACT

Contemporary psycholinguistic models suggest that there may be two possible routes in phonetic encoding: a 'direct' route which uses stored syllabic units, and an 'indirect' route which relies on the on-line assembly of sub-syllabic units. The computationally more efficient direct route is likely to be used for high frequency words, whereas the indirect route is most likely to be used for novel or low frequency words.

This paper presents some acoustic evidence that suggests that there may be dual routes operating in phonetic encoding. The data reported suggest that a group of normal speakers may be employing different routes in the phonetic encoding of high and low frequency words elicited via a repetition task. The evidence is presented and discussed within the framework of a dual-route hypothesis, and in light of other acoustic evidence reported in the literature.

## 1. INTRODUCTION

There has been some debate about whether articulatory sequences are calculated anew each time or retrieved from storage as and when necessary during speech production [1]. Some psycholinguistic models suggest that the speech segment has a critical role in the encoding of spoken output [2, 3, 4]. In these models, speech encoding is viewed as a process of segment-by-segment access and then assembly of the syllable/word. Much of the empirical evidence for such theories rests on speech error data from normal speakers (e.g. segmental switches which convert 'car park' to 'par cark'). There is, however, an alternative model that suggests that there may be two routes that may be operating in phonetic encoding [5]. This model suggests that speakers have a mental syllabary in which frequently used syllables are stored [5]. This is the route that we will refer to as the 'direct route'.[1] A parallel 'indirect route',[2] which is used for novel or low frequency syllables, relies more heavily on on-line computation and the assembly of the syllable from sub-syllabic units.

The indirect route is also more likely to be used in careful speech scenarios such as delivering a lecture, when more conscious attention is being directed to speech. We would therefore suggest that the dual-route hypothesis might also explain some of the phonetic phenomena that have been observed in the study of speech that has been elicited in different ways. For example, consonant-vowel gestures in spontaneous-like speech have been found to show more coarticulation than corresponding consonant-vowel gestures in non-spontaneous speech [6, 7].

Some of the measures that have been used to gauge the employment of direct and indirect routes have included response latencies and the duration of utterances. Here, greater values for both the aforementioned measures could be interpreted as a sign of greater planning and encoding demanded by the 'indirect' route [5]. It has also been suggested that there may be other indices which can be used to determine the level of 'direct'/'indirect' route encoding and these may include higher syllable cohesion (or increased coarticulation or gestural overlap) for syllables generated using the 'direct' route [8].

This preliminary study examines whether different routes may be operating in the phonetic encoding of high and low frequency monosyllabic words, elicited via a repetition task. The effect of word frequency on measures of response latency, utterance durations, word durations, consonant-vowel frequency changes was assessed. The results of this preliminary study are discussed within the framework of the dual-route hypothesis of phonetic encoding.

## 2. METHODOLOGY

### 2.1. Subjects

The subjects who participated in the study were 6 women (mean age=52.8 years, age range 47 to 60 years). All six subjects worked in a tertiary education environment. All speakers were from Sheffield and were matched for accent. None of the speakers had any speech, language or hearing impairment.

### 2.2. Speech Material

The speech material used in the study consisted of three repetitions of a set of phonetically matched 10 high frequency (occurrence of more than 100 per million) and 10 low frequency (occurrence of less than 10 per million) words, which were selected from Thorndike and Lorge [9].

The high frequency words were: *mile*, *base*, *pound*, *bag*, *school*, *cold*, *foot*, *group*, *cup* and *car*. The low frequency words were: *bile*, *mace*, *mound*, *bog*, *stool*, *colt*, *soot*, *croup*, *cub*, and *tar*. This gave a total of thirty tokens for each of the high and low frequency word groups, which were than randomised into a single list. Subjects were instructed to repeat each word after the experimenter. They were also required to prefix each word with the definite article *the* (e.g. *the bile, the mile*). All sessions were recorded in a quiet room using a DAT recorder. The data reported here is only for the first repetition of the data sets.

---

[1] This is referred to as the 'indirect' route by Levelt and Wheeldon [5].

[2] This is referred to as the 'direct' route by Levelt and Wheeldon [5].

## 2.3. Analyses

All speech data were digitized (SR 10kHz) and analysed using a KAY Computerized Lab (CSL) Model 4300. Speech pressure waveforms, wideband FFT spectrograms and LPC analyses were used to obtain the acoustic measures. The measures that were taken included:

- response (or repetition) latencies - these were measured from the end of the experimenter's prompting utterance to the start of the participant's utterance;

- utterance durations - these were measured from the start to the end of the entire utterance (i.e. *The Mile*);

- word durations - these were measured from the start to the end of the word;

- consonant-vowel[3] (CV) coarticulation measures were taken for the bilabial high/low frequency pairs (*mile/bile*, *base/mace*, *pound/mound*) and the velar high/low frequency pairs (*cup/cub*, *cold/colt*, *group/croup*)[4]. These included measurements of the second formant frequency (F2) at the onset of the vowel and the temporal midpoint of the vowels. For the plosive contexts, the vowel onset was taken at the point of the release of the plosive and for the nasal contexts, this was taken at the onset of the vowel following cessation of the low frequency nasal murmur. For the *group/croup* word pair, the vowel onset was taken at the end of the retroflex phase of the alveolar approximant in the clusters [gɹ] and [kɹ]. This was determined by end of the lowered frequency (or 'dip') pattern of the third formant which characterises the alveolar approximant[ɹ].

- The F2 difference between the temporal midpoint of the vowel and the vowel onset was then calculated for both groups of data. The time intervals between the vowel onset and temporal midpoints were measured as CV timelag durations.

# 3. RESULTS

## 3.1. Temporal Measures

The means and standard deviation values for response latencies, utterance durations and word durations are given in Table 1 for the high and low frequency words. Paired t-tests for response latencies indicated significant differences between the high and low frequency words ($t(59)=-4.106$, $p<.001$), with longer response latencies being found for the low frequency words. Paired *t*-tests for both utterance and word durations indicated that although the data showed evidence of a mean trend of longer utterance and word durations for the low frequency words, these differences were not significant.

The mean and standard deviation values of the degree of F2 change and CV Timelag for the high-low frequency bilabial data are given Table 2. Paired t-tests for both the F2 change and CV Timelag data sets indicated no significant differences between the high and low frequency words.

| | Word Frequency | |
|---|---|---|
| | High | Low |
| Response Latency | 170.1 (82.1) | 226.0 (115.7) |
| Utterance Duration | 652.6 (94.7) | 669.0 (111.2) |
| Word Duration | 538.0 (104.6) | 555.4 (132.2) |

**Table 1**: Mean and standard deviation values for response latency, utterance duration, word duration (in milliseconds) by word frequency for all data.

| | Word Frequency | |
|---|---|---|
| | High | Low |
| F2 change (Hz) | 186.9 (314.0) | 145.7 (277.7) |
| CV Timelag (ms) | 123.3 (29.3) | 111.6 (14.5) |

**Table 2**: Mean and standard deviation values for the degree of F2 change from vowel onset to mid vowel (Hz), and CV timelag values (milliseconds) for the bilabial data.

| | Word Frequency | |
|---|---|---|
| | High | Low |
| F2 change (Hz) | -295.3 (317.1) | -366.7 (263.6) |
| CV Timelag (ms) | 125.8 (32.8) | 119.5 (28.1) |

**Table 3**: Mean and standard deviation values for the degree of F2 change from vowel onset to mid vowel (Hz), and CV Timelag values (milliseconds) for the velar data.

The mean and standard deviation values of the degree of F2 change and CV Timelag for the high-low frequency velar data are given in Table 3. These data are also illustrated in Figures 1 and 2 for the degree of F2 change and the F2 Timelag, respectively. Paired t-tests for the F2 change and CV Timelag parameters indicated no significant differences between the high and low frequency words. However, as indicated in Figure 1, there appeared to be a trend in the F2 data to suggest that the high frequency words displayed a lower degree of formant movement. This pattern could be interpreted as some evidence for a higher degree of cohesiveness in the CV gestures of the high frequency words. The high frequency velar subset also showed some relationship between the CV Timelag values and corresponding F2 changes. A Pearson's product moment correlation test showed a significant negative correlation between the CV Timelag values and corresponding values of F2 change for the high frequency data ($r$= -.621, $p$<.01). This negative correlation is illustrated by the scatter plot in Figure 3 where we see a general pattern of longer CV Timelag values being correlated with greater negative F2 changes. The pattern of negative changes can be explained by the negative excursions

[3] The 'vowels' in these contexts were diphthongs.
[4] These word pairs were chosen because they were most closely phonetically matched in terms of articulatory constraints.

of the second formant into the back vowel/diphthong[5] contexts of *cup/cub* ([kʊp], [kʊb])[6], *cold/colt* ([koʊld], [koʊlt]) and *group/croup* ([gɹuːp/, [kɹuːp]). These back vowels tend to be characterised by a low second formant frequency.

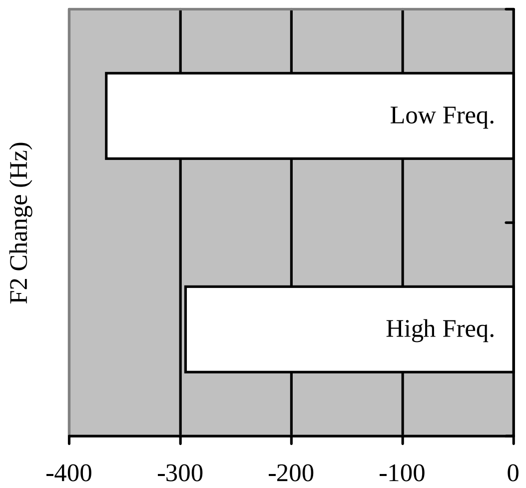


**Figure 1:** Mean and standard deviation values for the degree of F2 change (Hz) for the velar subset data given by word frequency.

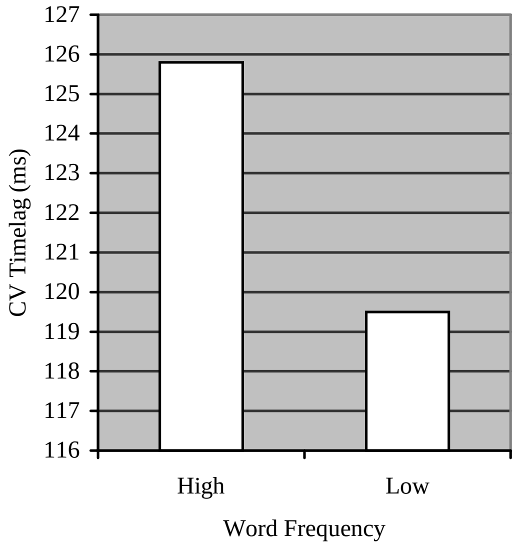


**Figure 2:** Mean and standard deviation values for the CV Timelag values (milliseconds) for the velar subset data given by word frequency.

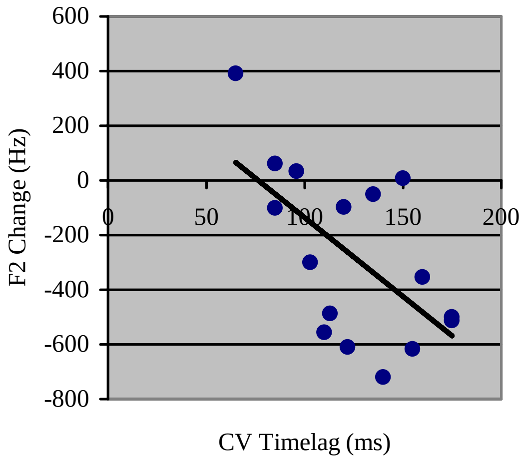


**Figure 3:** A scatter plot showing the correlation between CV timelag values (ms) and corresponding changes in F2 frequency (Hz) for the high frequency velar subset.

## 4. DISCUSSION

There is some evidence from the data of the subjects in this study which suggests that there may be dual routes operating in phonetic encoding. In the main, this evidence comes from the response latencies of the low frequency words being significantly longer than those of the high frequency words. Although no significant differences were found between the utterance and word durations of the high and low frequency words, the data displayed a trend of the low frequency words having longer word and utterance durations than those of the high frequency words.

The fact that we do not observe any significant differences between the high and low frequency words for either the F2 change values, or the CV Timelag duration values, could be interpreted in a number of ways. Firstly, the size of both the bilabial and velar data samples is small, and is therefore not sufficient to form the basis of any conclusive comments. Secondly, inter-speaker variability may be an additional factor contributing to the lack of significant differences in these two parameters.

The data reported here are preliminary but they are suggestive that the dual route hypothesis is useful in guiding further exploration of speech encoding. The data of the subjects reported here showed significantly longer response latencies to low frequency words than to the phonetically-matched high frequency cognate. The interpretation of this finding is not unambiguous. It might be attributable to linguistic factors such as slower lexical access to a low frequency word, or, the claim made by the dual route theory, that low frequency words are not stored as wholes, and therefore have to be assembled. By this claim, slower response latencies reflect greater planning time. This evidence might support the hypotheses of dual route theory. Further data collection would be useful to establish whether these trends reach significance.

---

[5] In this particular context we are referring to the off-glide of the diphthong

[6] All phonetic transcriptions are given in a broad notation.

The durational measures appeared to be in the main, more sensitive to differences of word frequency status than the changes in F2 frequency measures. The coarticulation measures were, however, based on a very small subset of observations and an expansion of the data set is required before firm conclusions can be drawn on the value of this measure in relation to the dual route hypothesis.

The debate over the merits of different models of speech encoding has largely been informed by data drawn from normal speakers. In other areas of psycholinguistic enquiry, for example, lexical processing in both speech and orthography, and in other domains of information processing, such as memory, there has been much greater use of data drawn from disorders. The convergence of evidence from neuropsychology and from studies of normal processing is indicative of a robust theoretical conceptualisation and models with wide predictive value [10]. The current study is a preliminary one and the trends within the data might be raised to significance with additional observations. A difficulty in analysis of normal performances for evidence of underlying processing mechanisms is that normal performance is 'seamless'. The surface behaviours may give little insight into the underlying processing systems and, in relation to the dual-route speech encoding hypothesis, products of different encoding routes may, on the surface, be only minimally different. Differences between data sets are therefore likely to be subtle and potentially capable of being obscured by outlier values such as individual differences between speakers in degree of habitual coarticulation. Similarly, differences in experimental paradigms may result in abolition of subtle trends in performance and a failure to replicate the results of studies across different laboratories.

The disorder which has most apparent relevance to theories of speech encoding is that of apraxia of speech (AOS). AOS is regarded as a disorder of phonetic planning which occurs post-lexical access and that is distinct from lower level failures of phonetic implementation (dysarthria). It is remarkable how little the behaviours of AOS patients [11, 12, 13, 14, 15, 16] have been used to inform theories of speech encoding, particularly in view of the extent to which other categories of disorder have had major influences on processing theories. Convergence of evidence from studies of normal speakers, albeit displaying subtle trends, with more robust evidence from speakers with AOS has the potential to develop more powerful theories of speech encoding.

## 5. ACKNOWLEDGMENTS

The authors wish to thank the participants, and our appreciation is also extended to H. Luff for contributions to earlier phases of the investigation.

## 6. REFERENCES

1. Keller, E. "The cortical representations of motor processes of speech", In: Keller, E. and Gopnik, M. (Eds.): *Motor and Sensory Processes of Language*. Hillsdale New Jersey, 1987.

2. Dell, G. "The retrieval of phonological forms in production: tests of prediction from a connectionist model", *Journal of Memory and Language*, 27, 1988, 124-142.

3. Shattuck-Hufnagel, S. "The role of word-onset consonants in speech production planning: New evidence from speech error patterns", In: Keller, E. and Gopnik, M. (Eds.) *Motor and Sensory Processes of Language*, Hillsdale New Jersey, 1987.

4. Shattuck-Hufnagel, S. "The role of word structure in segmental serial ordering", In: Levelt, W. J. M. (Ed.) *Lexical Access in Speech Production*, Oxford: Blackwell, 1993.

5. Levelt, W. J. M. and Wheeldon, L. "Do speakers have access to a mental syllabary?", *Cognition*, 50, 1994, 239-269.

6. Duez, D. "Second-formant locus patterns: An investigation of spontaneous French speech", *Speech Communication*, 11, 1992, 417-427.

7. Krull, D. "Second formant locus pattern and consonant-vowel coarticulation in spontaneous speech", Phonetic Experimental Research at the Institute of Linguistics, University of Stockholm, *Perlius* , 10, 1989, 87-108.

8. Whiteside, S. P. and Varley, R. A. "A new conceptualisation of apraxia of speech: a synthesis of evidence", *Cortex*, 34, 1998, 221-231.

9. Thorndike, E. L. and Lorge, I. *The Teacher's Wordbook of 30,000 Words*, New York: Teachers College Columbia University, 1944.

10. Shallice, T. *From Neuropsychology to Mental Structure*, Cambridge: C.U.P., 1988.

11. Kent, R. D. and Rosenbek, J. C. "Acoustic patterns of apraxia of speech", *Journal of Speech and Hearing Research*, 26, 1983, 231-249.

12. Kent, R. D. and Rosenbek, J. C. "Prosodic and neurologic lesion", *Brain and Language*, 15, 1982, 259-291.

13. Hardcastle, W. J. "Electropalatographic study of articulation disorders in verbal dyspraxia", In: Ryalls, J. H. (Ed.), *Phonetic Approaches to Speech Production in Aphasia and Related Disorders*, College-Hill Press, Boston, 1987.

14. Zeigler, W. and Von Cramon, D. "Disturbed coarticulation in apraxia of speech: acoustic evidence", *Brain and Language*, 29, 1986, 34-47.

15. Varley, R. A. and Whiteside, S. P. "Voicing in severe apraxia of speech: perceptual and acoustic analysis of a single case", *Journal of Neurolinguistics*, 11(3), 1998, 259-273.

16. Lebrun, Y. Apraxia of speech: a Critical Review. *Journal of Neurolinguistics*, 5 (4), 1990, 379-406.